

KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschung

Hollstein, Betina; Miller, Bernhard; Siegers, Pascal; Wolf, Christof

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Hollstein, B., Miller, B., Siegers, P., & Wolf, C. (2021). KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschung. *Bausteine Forschungsdatenmanagement*, 2, 10-22. <https://doi.org/10.17192/bfdm.2021.2.8330>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschungⁱ

Betina Hollstein

Bernhard Miller
Christof Wolf

Pascal Siegers

2021

Zitiervorschlag

Hollstein, Betina, Bernhard Miller, Pascal Siegers und Christof Wolf. 2021. KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschung *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2021: S. 10-22. DOI: [10.17192/bfdm.2021.2.8330](https://doi.org/10.17192/bfdm.2021.2.8330).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱBetina Hollstein (ORCID: [0000-0003-0219-6265](https://orcid.org/0000-0003-0219-6265)), Bernhard Miller (ORCID: [0000-0002-4385-7245](https://orcid.org/0000-0002-4385-7245)), Pascal Siegers (ORCID: [0000-0001-7899-6045](https://orcid.org/0000-0001-7899-6045)), Christof Wolf (ORCID: [0000-0002-9364-9524](https://orcid.org/0000-0002-9364-9524))

SWD - Strengthen, Widen, Deepen. so versteht das Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften seine Aufgabe innerhalb der Nationalen Forschungsdateninfrastruktur: die bestehende Infrastruktur für Forschungsdatenmanagement in unseren Disziplinen auszubauen und zu integrieren. Um gesellschaftlich relevante, insbesondere sensible Daten für die Forschung verfügbar zu machen, baut KonsortSWD auf bewährten, wissenschaftsnahen Strukturen auf. Wir integrieren die bestehenden Elemente stärker und erschließen systematisch neue Daten, z.B. qualitative Forschungsdaten. Eine zentrale Rolle spielt dabei der Rat für Sozial- und Wirtschaftsdaten, der als gewähltes Gremium die Anliegen der Fachcommunities artikuliert und die zielgerichtete Weiterentwicklung der Infrastruktur forciert.

1 Stärke aus Dezentralität

Mit dem Konsortium für die Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften (KonsortSWD)¹ wird innerhalb der Nationalen Forschungsdateninfrastruktur (NFDI) ein bereits erfolgreich etabliertes Kooperationsnetzwerk aus Forschungsdatenzentren und des RatSWD zu einer integrierten Dateninfrastruktur weiterentwickelt. Die Stärke dieser Dateninfrastruktur ist das geteilte Verständnis, dass viele Daten nur dann für die (Sekundär-)Nutzung bereitgestellt werden können, wenn sie bei den Datenanbieter:innen verbleiben. Der Grund ist, dass die Datenanbieter:innen aufgrund ihrer Expertise zugleich die höchstmögliche Qualität bei der Erschließung und Aufbereitung dieser Daten sichern können. Daraus entsteht eine dezentrale und deshalb in hohem Maße leistungsfähige Infrastruktur.

Dass Dezentralität gleichzeitig auch Effizienz bedeutet, hat mit der Vielfalt von Datentypen, vor allem aber mit der Vielfalt an rechtlichen und forschungsethischen Einschränkungen zu tun, die für die Beforschung von Daten in den Disziplinen von KonsortSWD bestehen. Diese lassen sich dann am besten im Sinne der Forschung handhaben, wenn ein hohes Maß an Expertise direkt „bei“ den Daten vorhanden ist. Versinnbildlicht wird diese hochgradig arbeitsteilige und dennoch zu stark koordinierten Lösungen fähige Struktur durch die Ameise, die sich KonsortSWD als Logo ausgewählt hat (vgl. Abbildung 1). Weil es schon seit langem das inhärente Ziel dieser Infrastruktur ist, Daten zu Forschungszwecken aufzubereiten und bereitzustellen, ist sie auch ein guter Ausgangspunkt für eine weitere Integration im Rahmen der NFDI und der European Open Science Cloud (EOSC).

KonsortSWD gründet wesentlich auf der Arbeit des Rates für Sozial- und Wirtschaftsdaten (RatSWD), der 2004 gegründet wurde.² Neben seiner maßgeblichen Rolle bei

¹KonsortSWD wird im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert - Projektnummer: 442494171.

²Nähere Informationen zu KonsortSWD unter <https://www.konsortswd.de>; dort findet sich auch eine stets aktuelle Übersicht über die Forschungsdatenzentren. Die Autor:innen danken Margit Bäck und Janina Jablonski für die Unterstützung bei der Erstellung des Manuskriptes.

der Beratung der Politik für forschungsfreundliche Zugänge zu administrativen Daten hat der RatSWD vor allem dem Modell der Forschungsdatenzentren (FDZ) zum Erfolg verholfen. FDZ stellen heute Daten deutlich über die namensgebenden Sozial- und Wirtschaftswissenschaften hinaus zur Verfügung.

Den frühen Start hin zu einem professionellen und forschungsfreundlichen Umgang mit Forschungsdaten verdanken die in KonsortSWD vertretenen Disziplinen der "Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik" (KVI)³. Ähnlich wie der Rat für Informationsinfrastrukturen⁴ offenbarte die Bestandsaufnahme der KVI im Jahr 2001 Defizite beim Zugang der Forschung zu hochrelevanten Daten, vor allem jenen, die von staatlichen Stellen erhoben wurden. 2004 folgte die Gründung des RatSWD als Beratungsgremium der Bundesregierung. Ziel war die Verbesserung des Zugangs der Wissenschaft zu Daten, die aus Gründen des Datenschutzes oder der Vertraulichkeit bis dahin nicht zu Forschungszwecken allgemein verfügbar waren. Als Grundlage für die später entstehende Infrastruktur etablierte der RatSWD einen strukturierten Dialog zwischen Datenproduzent:innen und Datennutzer:innen, um wechselseitiges Verständnis zu schaffen. Die Zusammensetzung des RatSWD spiegelt dieses Ziel wider.⁵

Seit seiner Gründung vernetzt der RatSWD eine stetig wachsende Infrastruktur, die der Vielfalt der Gesellschaftsforschung mit ihren unterschiedlichen Datenbedarfen Rechnung trägt. Ein FDZ archiviert Daten und macht diese über verschiedene Zugangswege unter Einhaltung des Datenschutzes für wissenschaftliche Zwecke zugänglich⁶. So konnten und können zahlreiche Datenbestände für die Forschung erschlossen werden, die vormals nicht zur Verfügung standen (z.B. Daten aus der amtlichen Statistik, den Sozialversicherungen oder in naher Zukunft auch aus dem Ausländerzentralregister). Dadurch, dass die Daten in den Organisationen verbleiben und nicht an ein Repositorium übertragen werden, können auch datenschutzrechtlich sensible Merkmale für die Forschung aufbereitet und – mit einem angemessenen Schutzkonzept – der Wissenschaft zugänglich gemacht werden. Die Dezentralität der RatSWD Dateninfrastruktur entspringt also dem Interesse der Wissenschaft. Die durch den RatSWD verantwortete Akkreditierung der nunmehr 38 FDZ sichert trotz unterschiedlichster Ausgangsbedingungen bei Datenarten, Finanzierung und rechtlichen Grundlagen einen qualitätsgesicherten Zugang von Forschenden zu diesen Daten.⁷ Die FDZ bieten Daten aus

³Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik, *Wege zu einer besseren informationellen Infrastruktur* (Baden-Baden: Nomos Verlagsgesellschaft, 2001).

⁴Rat für Informationsinfrastrukturen, *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland* (Göttingen: Rat für Informationsinfrastrukturen, 2019), [urn:nbn:de:101:1-201606229098](https://nbn-resolving.org/urn:nbn:de:101:1-201606229098)

⁵Zehn Ratsmitglieder werden von Vertreter:innen der amtlichen und wissenschaftlichen Datenproduzent:innen nominiert. Zehn weitere werden auf Vorschlag der Fachgesellschaften von Wissenschaftler:innen gewählt. An der letzten Wahl beteiligten sich im Winter 2020 fast 3.500 Forschende.

⁶Rat für Sozial- und Wirtschaftsdaten (RatSWD), *Tätigkeitsbericht 2019 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ)* (Berlin 2020), [doi:10.17620/02671.56](https://doi.org/10.17620/02671.56)

⁷Wissenschaftsrat, *Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissen-*

Wissenschaft, Verwaltung und Zivilgesellschaft für eine Nachnutzung an und haben eine breite Nutzendenbasis. Im Jahr 2019 wurden 4.371 Datenfiles angeboten und von mehr als 55.000 Forschenden verwendet.⁸ Die Bestände der FDZ bilden das gesamte Themenspektrum der Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften ab: Daten aus standardisierten Umfragen und amtlichen Registern, Audioaufzeichnungen, Textdateien und Transkripte, Videoaufzeichnungen, Geodaten und vieles mehr.

Daneben hat sich in den vergangenen Jahren das wissenschaftspolitische Profil des RatSWD gewandelt. Neben seinem aktiven Einsatz für einen verbesserten Datenzugang und verbesserte Datenqualität hat die Begleitung laufender Gesetzgebungsprozesse (wie z.B. der EU-Datenschutzgrundverordnung) an Bedeutung gewonnen. Vermehrt werden Implikationen neuer Technologien in der Datenerhebung (z.B. BigData, Sensor-Daten) diskutiert und Empfehlungen für die Forschung formuliert. Gleichzeitig bleibt der Rat Ansprechpartner von Ministerien, Wissenschaftler:innen und Verwaltung in wissenschaftspolitischen Fragen zu (Forschungs-)Daten.

Auch über die FDZ und die Aktivitäten des RatSWD hinaus besteht in den Disziplinen ein großes Interesse an Daten aus anderen Wissenschaftsbereichen und an einem weiter vereinfachten Zugang zu den existierenden Datenbeständen. SWD steht in der NFDI somit für die Stärkung (strengthen), die Erweiterung (widen) sowie die Vertiefung (deepen) der Zusammenarbeit zwischen Forschenden und Infrastrukturanbietenden in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften.

2 Strengthen, Widen, Deepen: Ziele und Struktur von KonsortSWD

KonsortSWD setzt sich für möglichst leicht nutzbare, qualitativ hochwertige Daten für die Gesellschaftsforschung ein. Die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften eröffnen vielfältige Perspektiven und entwickeln wesentliche Ansätze, um Antworten auf große aktuelle gesellschaftliche Herausforderungen zu finden. Unsere Mission ist es, die Forschungsdateninfrastruktur für die Gesellschaftsforschung zu stärken, zu erweitern und zu vertiefen. Dies geschieht nutzungsorientiert und unter Berücksichtigung der Bedürfnisse der Forschenden in unseren Disziplinen. Unser Forschungsdatenmanagement wird daher so angelegt sein, dass es Forschende und Forschungsdatenzentren technisch und inhaltlich bei der Sicherung und Nachnutzung bestehender und neuer sensibler ebenso wie nicht sensibler Daten unterstützt. Dies erfolgt in Übereinstimmung mit den FAIR-Prinzipien⁹ und unter Sicherstellung ethischer Standards und rechtlicher Vorgaben des Datenschutzes.

schaften (Drs. 10465-11) (Berlin: Wissenschaftsrat, 2011), 81f.

⁸RatSWD, *Tätigkeitsbericht 2019 der vom RatSWD akkreditierten FDZ*.

⁹Noemi Betancort Cabrera et al., "White Paper on implementing the FAIR principles for data in the Social, Behavioural, and Economic Sciences" *RatSWD Working Papers* 274 (2020), doi:10.17620/02671.60.

Eine zentrale Rolle kommt dabei dem beständig wachsenden Netzwerk von derzeit 38 Forschungsdatenzentren zu. Denn FDZ kuratieren und archivieren Forschungsdaten und machen diese über verschiedene Zugangswege unter Einhaltung des Datenschutzes für wissenschaftliche Zwecke zugänglich. Sie zeichnen sich zumeist durch einen fachlichen Schwerpunkt (z.B. Arbeit, Bildung, Gesundheit) verbunden mit spezifischer Kompetenz für Daten aus bestimmten Erhebungsverfahren (z.B. Umfragedaten, Beobachtungsdaten, Registerdaten) und oder bestimmten digitalen Objekttypen (z.B. Videos, Tabellendaten, Textkorpora) aus. Die FDZ verbinden damit spezifische inhaltliche und methodische Kompetenz, die sie für eigene Forschung und die Beratung von Datennutzer:innen einsetzen. Dieses Netzwerk bildet damit eine einzigartige Grundlage für qualitativ hochwertige Daten für die Forschung. Die Akkreditierung durch den RatSWD sichert Mindeststandards für alle FDZ und ermöglicht ein leistungsfähiges Monitoring. Die von KonsortSWD entwickelten Innovationen im Datenmanagement werden – vermittelt über die FDZ – der wissenschaftlichen Nutzung zu Gute kommen. Neben ihrer Verantwortung für den Datenzugang sind die FDZ also auch Multiplikatoren für die Angebote von KonsortSWD.

KonsortSWD hat sich bewusst dafür entschieden, bewährte Strukturen für die Zusammenarbeit, die Standardsetzung und die politische Beratung zu Forschungsdaten in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften in das Arbeitsprogramm des Konsortiums zu integrieren. Aus diesem Grund bleibt der RatSWD, dessen wissenschaftliche Mitglieder von 15 Fachgesellschaften nominiert und von Forschenden direkt gewählt werden, als unabhängiges und qualitätssicherndes Gremium für unsere Disziplinen als Teil von KonsortSWD erhalten.¹⁰ Die große Bekanntheit des RatSWD trägt dazu bei, die Verfügbarkeit von Daten zu verbessern. Seine Akkreditierung gilt als Ausweis für einen transparenten Datenzugang. Wissenschaftliche Beiräte und Evaluationskommissionen empfehlen die Akkreditierung deshalb vermehrt im Rahmen von internen oder externen Audits. Auch die Datenstrategie der Bundesregierung zielt auf die Einrichtung neuer Forschungsdatenzentren in den Geschäftsbereichen aller Ministerien und die Stärkung der Forschungsdatenzentren in der Verantwortung der Länder.¹¹ Über den Datenzugang hinaus, engagiert sich der RatSWD für die Sicherstellung der ethischen und anderer Rahmenbedingungen der Datennutzung

¹⁰Die derzeit 15 für den RatSWD nominierungsberechtigten Fachgesellschaften sind: Akademie für Soziologie (AS), Deutsche Gesellschaft für Erziehungswissenschaft (DGfE), Deutsche Gesellschaft für Gesundheitsökonomie (DGGÖ), Deutsche Gesellschaft für Medizinische Psychologie (DGMP), Deutsche Gesellschaft für Politikwissenschaft (DGfP), Deutsche Gesellschaft für Psychologie (DGPs), Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPuK), Deutsche Gesellschaft für Sozial- und Kulturanthropologie (DGSKA), Deutsche Gesellschaft für Soziologie (DGS), Deutsche Gesellschaft für Volkskunde (DGV), Deutsche Statistische Gesellschaft (DStatG), Deutsche Vereinigung für Politikwissenschaft (DVPW), Gesellschaft für empirische Bildungsforschung (GEBF), Verband der Hochschullehrer für Betriebswirtschaft (VHB) sowie Verein für Socialpolitik (VfS). Eine aktuelle Liste der Mitglieder s. <https://www.konsortswd.de/ratswd/der-ratswd/mitglieder/>.

¹¹Bundesregierung. *Datenstrategie der Bundesregierung. Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum - Kabinettdfassung*, (Berlin: 2021), <https://www.bundesregierung.de/publikationen>, abgerufen am 13. Mai 2021.

im Sinne der Forschung. Die Akkreditierung von FDZ wird auch in Zukunft durch den RatSWD erfolgen.

Neben der oben beschriebenen wissenschaftspolitischen Tätigkeit wird der Rat für KonsortSWD in Zukunft noch weitere Funktionen übernehmen. So kann er helfen, Angebote des Konsortiums in den Fachgemeinschaften bekannt zu machen und Bedarfe der Communities im Konsortium zu vertreten. Dies geschieht unter anderem dadurch, dass der Beirat von KonsortSWD aus Mitgliedern des RatSWD besetzt wird.

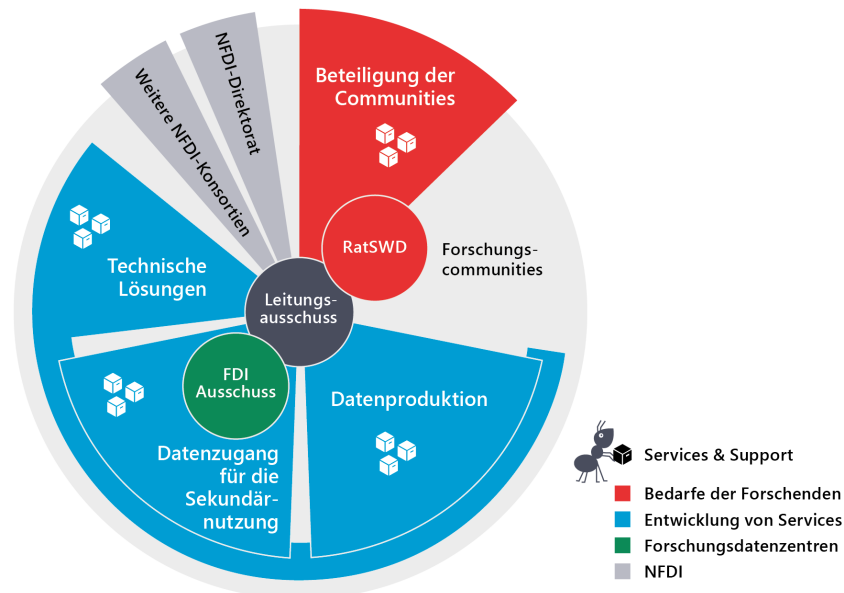


Abbildung 1: Struktur von KonsortSWD.

Abbildung 1 zeigt dieses Zusammenspiel von RatSWD, den inhaltlichen Arbeiten in den Task Areas des Konsortiums sowie mit dem Ausschuss für Forschungsdateninfrastruktur. Dabei weitet KonsortSWD die Einbindung der Forschenden auch noch über der Rat hinaus aus, z.B. über die direkte Interaktion auch mit den Fachgemeinschaften, die keine Vertreter:innen in den RatSWD entsenden.

Um in Zusammenarbeit mit KonsortSWD die FDZ-Infrastruktur weiter auszubauen und noch besser an die Bedarfe der Nutzenden anzupassen, arbeitet das Konsortium eng mit dem Ausschuss für Forschungsdateninfrastruktur beim Rat für Sozial- und Wirtschaftsdaten zusammen. Um die Daten, die in den Forschungsdatenzentren bereitstehen, gut findbar aber auch technisch gut nachnutzbar zu machen, stellt KonsortSWD außerdem seine Expertise für technische Lösungen bereit.

Für die **Stärkung** (strengthen) der Infrastruktur wird KonsortSWD vor allem die FAIRness von Daten und Metadaten ausbauen. Diese kann auf den Arbeiten der Data Documentation Initiative (DDI) aufbauen, einem international verwendeten Metadatenstan-

dard, der maßgeblich von der KonsortSWD-Community entwickelt wurde und wird.¹² Es gilt, diesen Standard zur Beschreibung von Daten, die durch Umfragen und andere Beobachtungsmethoden erzeugt werden, auch zur Verwendung für die vielfältigen weiteren Datentypen anzupassen und seine Verwendung zu vereinfachen. Besonders zu nennen sind hier unstrukturierte Daten z.B. in der qualitativen Sozialforschung oder, allgemeiner, im Bereich von Textdaten (auch aus den sozialen Medien). Insgesamt ist vor allem die Stärkung der Interoperabilität von Forschungsdaten ein Ziel, das durch die Arbeit von KonsortSWD greifbarer werden soll, indem die Granularität und Qualität der Metadaten gesteigert werden.

Erweiterung (widen) bedeutet, dass wir neue (sensible) Datentypen in die Infrastruktur miteinbeziehen (z.B. strukturierte Interview-, Beobachtungs- und Textdaten¹³¹⁴, Daten, die mit Web-Scraping-Technologien gesammelt wurden, oder Daten aus dem öffentlichen Gesundheitswesen sowie Unternehmensdaten).¹⁵ Durch die Ausweitung der Reichweite der FDZ profitieren neue Communities vom Datenzugang und von der im KonsortSWD-Netzwerk verfügbaren Expertise im Forschungsdatenmanagement. Ein Beispiel sind die sozial- und kulturanthropologischen Fachcommunities, die mit der Ausweitung auf unstrukturierte qualitative Daten für ihre Forschungsdaten erstmals eine Unterstützungsstruktur erhalten. Zudem eröffnen sich neue und innovative Möglichkeiten, Daten auch über die Grenzen von Disziplinen hinweg zu verknüpfen, z.B. Satellitendaten mit amtlichen Daten und Umfragedaten zu sozialen Rahmenbedingungen in individuellen Lebensräumen.

Eine **vertiefte** Zusammenarbeit (deepen) zwischen den FDZ zielt darauf, die aktuelle Fragmentierung der Datenlandschaft weiter zu überwinden. Als Grundlage für diese vertiefende Zusammenarbeit haben die Vertreter:innen der FDZ seit 2017 auf Basis ihrer täglichen Erfahrungen in der Zusammenarbeit mit ihren Nutzer:innen strategische Handlungsfelder identifiziert: Hierzu zählen die Harmonisierung von Standards und Prozessen in den FDZ sowie ein Netzwerk von "Single Points of Access", um den Datenzugang zu erleichtern und die Transparenz erhöhen. Ein gemeinsames Trainingsprogramm für FDZ-Mitarbeiter:innen und Nachwuchswissenschaftler:innen wird zusätzlich die FDM-Kompetenzen vertiefen und für die voranschreitende Digitalisierung der Forschung rüsten. Für die Forschenden erleichtert diese Vertiefung der Zusammenarbeit zwischen den FDZ und die Vereinheitlichung von Standards und Prozessen den Datenzugang.

¹²Claus-Peter Klas und Oliver Hopt, "An Operationalized DDI Infrastructure to Document, Publish, Preserve and Search Social Science Research Data," in *International Conference on Theory and Practice of Digital Libraries*, Eva Méndez, ed, (Cham: Springer, 2018), 94-99.

¹³Betina Hollstein, "Qualitative Approaches," in *The SAGE Handbook of Social Network Analysis*, John Scott und Peter J. Carrington, ed, (London/New Delhi: Sage Publications, 2011), 404-417.

¹⁴Betina Hollstein und Jörg Strübing, "Archivierung und Zugang zu Qualitativen Daten," *RatSWD Working Papers* 267 (2018), doi: [10.17620/02671.35](https://doi.org/10.17620/02671.35).

¹⁵Die FDZ decken schon heute eine große Bandbreite ab. Dies wird auch deutlich darin, dass FDZ in mehreren NFDI-Konsortien aktiv sind. So ist z.B. das FDZ GePaRD (German Pharmacoepidemiological Research Database) Teil von NFDI4Health, die FDZ der Universität Mannheim und des ZEW von BERD4@NFDI und das Archiv für Gesprochenes Deutsch (AGD) von Text+.

Bei den beschriebenen Vorhaben für die Stärkung, Erweiterung und Vertiefung der Infrastruktur spielt der RatSWD eine zentrale Rolle. Neben seiner Rolle als Qualitätssicherer für die FDZ artikuliert der Rat sowohl herausgehobene Expertise als auch die Wünsche der Fachcommunities, auch gegenüber der Politik. Dies betrifft nicht zuletzt auch die rechtlichen Rahmenbedingungen der Datenproduktion und des Datenzugangs, die für eine forschungsfreundliche Dateninfrastruktur eine zentrale Rolle spielen.

3 Wichtige Beiträge von KonsortSWD zum Forschungsdatenmanagement

Wichtige Beiträge zum Forschungsdatenmanagement durch KonsortSWD sind die Stärkung der Bereitstellung FAIRer Forschungsdaten, die Erweiterung um unstrukturierte, insbesondere qualitative Daten, Kompetenzentwicklung sowie die Vertiefung der Zusammenarbeit innerhalb der Dateninfrastruktur.

Die **FAIR-Prinzipien** sind Leitlinien für den Aufbau der NFDI sowie speziell auch für KonsortSWD. Ihre Implementierung ist allerdings keineswegs, wie manchmal suggeriert wird, ein vorwiegend technischer Prozess. Vielmehr setzt sie ein hohes Maß an Standardisierung und Qualität im Daten- und Metadatenmanagement der Forschungsdatenzentren und damit substanzwissenschaftliche Expertise voraus. In einer dezentralen Struktur sind die Kompetenzen für das Forschungsdatenmanagement ungleich verteilt, weil die Mitarbeiter:innen der FDZ in erster Linie Fachwissenschaftler:innen mit einer großen Expertise für die Daten und die Methoden der Datengewinnung sind. Kompetenzen zu Metadatenstandards, im Metadatenmanagement und für die digitale Langzeitsicherung der Forschungsdaten werden meistens über die Tätigkeit im FDZ erworben. KonsortSWD unterstützt bestehende und neue FDZ bei der Professionalisierung des Datenmanagements. Ein Kernelement ist die Zertifizierung der FDZ als vertrauenswürdige digitale Repositorien (z.B. über das CoreTrustSeal), weil so grundlegende Prozesse wie die Langzeitarchivierung von Daten und Metadaten formalisiert und damit Auffindbarkeit und Zugänglichkeit gestärkt werden.

Für die Nutzenden noch relevanter ist die Verbesserung der Auffindbarkeit der Daten. In der Praxis hat sich gezeigt, dass durch die Verwendung von persistenten Identifikatoren und detaillierten Studien- und Variablenbeschreibungen Daten zwar auffindbar werden. Die Suche in Spezialportalen bleibt jedoch unübersichtlich, weil nur Ausschnitte aus den Metadaten (z.B. DublinCore) indexiert und damit die Inhalte aus den Daten nicht ausreichend abgebildet werden (können). Die Zukunft der Datensuche liegt unserer Ansicht nach nicht in dezidierten Suchportalen, wie sie derzeit in vielen Bereichen entwickelt werden, sondern in einer Annotation der Metadaten nach Standards des World Wide Web Consortiums wie z.B. schema.org. Damit werden die Daten aus den FDZ-Portalen und -Katalogen unmittelbar für die allgemeinen Suchmaschinen im Web

indizierbar und so für die Nutzenden sichtbar. Gestützt wird dies durch die Vergabe von persistenten Identifikatoren auf Ebene von Attributen unterhalb ganzer Datensätzen (in-line objects). So werden auch granulare Metadaten der FDZ direkt aus generischen Suchindizes heraus erreichbar und damit einfacher nachnutzbar. Dazu wird KonsortSWD den Zugang zu Daten und Metadaten innerhalb der eigenen Infrastruktur über Programmierschnittstellen (API) ausbauen.

Außerdem sieht KonsortSWD auch die Kompetenzentwicklung zum **Daten- und Metadatenmanagement** in den FDZ und Forschungscommunities als zentralen Baustein der Forschungsdateninfrastruktur.¹⁶ Entwickelt wird daher ein Trainingsangebot mit besonderem Fokus auf digitalen Kompetenzen für das Datenmanagement. KonsortSWD beteiligt sich daher auch an "Data Train", einer Lehrreihe zu Datenmanagement und Data Science der University Bremen Research Alliance, die gemeinsam mit Akteur:innen aus vier NFDI-Konsortien entwickelt wurde.¹⁷

Daneben ist eine wesentliche strategische Erweiterung der Forschungsdateninfrastruktur die Inklusion von - im Vergleich zu quantitativen Forschungsdaten - gering strukturierten **qualitativen Forschungsdaten**. Qualitative oder ethnographische Methoden und Forschungsdaten machen einen bedeutenden Teil der sozialwissenschaftlichen Forschung aus. In den ethnologischen Fächern ist die Forschung sogar fast ausschließlich qualitativ ausgerichtet. Qualitative Methoden und Forschungsdaten zeichnen sich durch gegenstandsbezogene Methodenvielfalt aus, wobei in den Forschungsprojekten vielfach unterschiedliche Methoden (Interviews, Beobachtungen, Dokumentenanalysen etc.) kombiniert und Daten in sehr unterschiedlichen Formaten und Medienformen gespeichert werden (bspw. als Textkorpora vorliegende Interviewtranskripte, Feldnotizen und Beobachtungsprotokolle sowie Bild-, Video- und Tondateien). Diese heterogenen, komplexen und zumeist personenbezogenen Daten standen bislang nur in Ausnahmefällen für die Sekundärnutzung zur Verfügung. Aufgrund dieser Komplexität und der hohen Schutzerfordernisse bei diesen Daten, aber auch weil für die Sekundärnutzung der Erhebungskontext besonders ausführlich dokumentiert werden muss, stellen diese Daten hohe Anforderungen an die Archivierung und Datenaufbereitung. Sie erfordern daher auch eine intensive Abstimmung zwischen Datenzentren und Fachcommunities, um im Forschungsprozess von Beginn an das FDM nachhaltig und nachnutzungsorientiert auszurichten.

Im Rahmen von KonsortSWD werden die bislang unverbundenen FDZ und Archive, die qualitative Primärforschungsdaten kuratieren, zu einer nachhaltigen vernetzten Struktur zusammengeführt. Damit betreten wir auch international Neuland: Etabliert wird eine föderierte, auf Nutzer:innen und Services ausgerichtete FDM-Infrastruktur für qualitative Primärforschungsdaten (QualidataNet). Sie ersetzt die bisherige heterogene,

¹⁶Rat für Informationsinfrastrukturen (RfII), *Digitale Kompetenzen - dringend gesucht! Empfehlungen zu Berufs- und Ausbildungsperspektiven für den Arbeitsmarkt Wissenschaft* (Göttingen: Rat für Informationsinfrastrukturen, 2019), [urn:nbn:de:101:1-2019080711032249706218](https://nbn-resolving.org/urn:nbn:de:101:1-2019080711032249706218).

¹⁷NFDI4BioDiversity, NFDI4Health, KonsortSWD und NFDI4Ing, <https://www.uni-bremen.de/research-alliance/forschungsdaten/data-train>.

fragmentierte und für Nutzende wenig transparente Archivlandschaft in diesem Bereich und bietet den Forschenden - Primärforscher:innen wie Sekundärnutzer:innen - eine zentrale Anlaufstelle für Information, Datenübergabe und Datenzugang. Diese föderierte Struktur wird sukzessive ausgebaut, indem neben den vom RatSWD akkreditierten FDZ weitere Bestände qualitativer Daten inkludiert werden. Aufbauend auf den Erfahrungen im Verbund Forschungsdaten Bildung ist die Umsetzung an zwei Prinzipien orientiert: Die Kuratierung und langfristige Sicherung erfolgen erstens dort, wo die thematische, methodische und technische Kompetenz für die Daten am größten ist. Zweitens entsprechen die Angebote für externe Datengebende den gleichen Qualitätsstandards wie sie in den FDZ für dort erhobene (interne) Daten gelten.¹⁸ Die Qualitätsstandards weisen dabei einen möglichst hohen Grad an Harmonisierung und Vergleichbarkeit auf.

Darüber hinaus wird in der föderierten Struktur des QualidataNet gemeinsam ein auf qualitative Daten und qualitative Forschung zugeschnittenes professionelles FDM-Portfolio entwickelt. Es wird Forschende effektiv und nachhaltig beim Forschungsdatenmanagement über den gesamten Forschungsprozess unterstützen (u.a. durch Anonymisierungstools und Handreichungen zu informierten Einwilligungen, Datenschutz und Daten-Kontextualisierung). Es genügt höchsten Schutzanforderungen bei gleichzeitig größtmöglichem Erhalt des Nachnutzungspotentials. Ein gemeinsames, an qualitative Daten und Methoden angepasstes Metadatenschema und kontrolliertes Vokabular helfen darüber hinaus, qualitative Daten FAIR zu präsentieren und international les- und findbar zu machen.

Wie das Beispiel der Dateninfrastruktur für unstrukturierte Daten zeigt, ist die **Vertiefung der Zusammenarbeit** zwischen den FDZ ein Schlüssel für die transparente Bereitstellung qualitativ hochwertiger Forschungsdaten für die Nutzenden. Auch an anderer Stelle ist es daher Ziel des Konsortiums, Zugangspunkte zu den Daten stärker zu vernetzen, um den Aufwand für die Nutzenden (z.B. bezogen auf Datensuche und Beurteilung des data fit, Datenaufbereitung und Datenabgabe) zu reduzieren. Dies gilt besonders für den Zugang zu sensiblen Daten an den Gastwissenschaftsarbeitsplätzen der FDZ. Diese bieten eine geschützte Arbeitsumgebung für den Zugang zu schwach anonymisierten oder pseudonymisierten Daten an, die andernfalls aus Gründen des Datenschutzes nicht zugänglich gemacht werden könnten. Bislang müssen Nutzende oft zu den jeweiligen FDZ reisen, um Zugang zu erhalten. In Zukunft soll ein bundesweites Netzwerk den Zugang zu sensiblen Daten erleichtern.

¹⁸Alexia Meyermann et al., "Der Verbund Forschungsdaten Bildung – Eine Forschungsdateninfrastruktur für die empirische Bildungsforschung," *RatSWD Working Papers* 266 (2017), doi: 10.17620/02671.27.

4 Herausforderungen

Die NFDI wurde gegründet, weil in allen Disziplinen große Herausforderungen für die FAIRe Nachnutzung von Forschungsdaten festgestellt wurden.^{19,20} Für KonsortSWD ergeben sich zentrale Herausforderungen aber nicht nur aus den Anforderungen FAIRer Forschungsdaten. Vielmehr auch, - und aus unserer Sicht oft bedeutender - aus den Kontextbedingungen für die Nutzung und das Teilen dieser Daten. Stichworte sind hier die Forschungsdatenethik ebenso wie die Einbindung der Communities, um sicherzustellen, dass die Forschungsdateninfrastruktur die Forschung bestmöglich unterstützt.

Alleinstellungsmerkmal der NFDI und damit auch eine zentrale Herausforderung ist der Aufbau dieser Infrastruktur für das Forschungsdatenmanagement aus den Fachgemeinschaften heraus.²¹ Während der RatSWD als gewählte Vertretung der Forschenden ein wesentliches Element der **Einbindung der Community in die Dateninfrastruktur** sicherstellt, will KonsortSWD die Forschenden auch darüber hinaus noch stärker und direkter in seine Arbeit einbinden. Dies nicht zuletzt mit dem Ziel eine Kultur des Datenteilens weiter zu fördern. So hat sich der RatSWD durch den NFDI-Prozess bereits noch weiter geöffnet (durch zwei zusätzliche Mitglieder und damit Raum für die Vertretung einer größeren Zahl an Fachgesellschaften). Zudem wird ein Community-Büro als ständige Anlaufstelle eingerichtet. Der Community-Beauftragte initiiert darüber hinaus Gespräche und bittet um Anregungen und Rückmeldungen aus den Fachgesellschaften und von Forschenden. Auf der Grundlage dieser Informationen entwickelt KonsortSWD seine Serviceangebote weiter. Durch diese starke Rückbindung an jene, die Mehrwert aus den bereitgestellten oder erschlossenen Forschungsdaten für ihre Arbeit ziehen, ist KonsortSWD eine lernende Struktur, die auch sicherstellt, dass auf Feedback rasch und adäquat reagiert wird.

Gleichzeitig kann die Einbindung der Forschenden aber nicht nur eine Holschuld der NFDI-Konsortien sein. Die Communities selbst haben auch eine Bringschuld: Sie müssen aktiv und explizit die Bedarfe für ein Forschungsdatenmanagement artikulieren, das ihnen hilft, ihre Forschungsfragen bestmöglich zu beantworten.

Eine Herausforderung in diesem Zusammenhang ist, das Teilen von Daten für die Sekundärnutzung in der Breite zu fördern. Hier können einzelne Sub-Communities, aber auch z. B. etablierte Forscher:innen, die in geringerem Maße auf viele Publikationen aus denselben Daten angewiesen sind, wertvolle Beispiele geben. Daher wird KonsortSWD im Rahmen eines wettbewerblichen Verfahrens Möglichkeiten auch für einzelne Forscher:innen schaffen, innovative Daten für die Sekundärnutzung in ihren

¹⁹RfII, *Leistung aus Vielfalt*.

²⁰Gemeinsame Wissenschaftskonferenz von Bund und Ländern, "Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)", 26. November 2018, <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf>.

²¹Petra Gehring und Stefan Lange, "Bitte nicht auf Autopilot schalten!," 02. Oktober 2020, <https://www.jmwiarda.de/2020/10/02/bitte-nicht-auf-autopilot-schalten/>.

Communities zu erschließen. Diese können dann mit finanzieller Unterstützung durch das Konsortium in den FDZ oder den angeschlossenen Repositorien bereitgestellt werden. Außerdem freuen wir uns auf Vorschläge aus unseren communities heraus, welche neuen Partnereinrichtungen mit neuen Diensten die Infrastruktur für das Forschungsdatenmanagement zusätzlich weiter entwickeln könnten.

Jenseits dieser Punkte ist in einer integrierten Forschungsdateninfrastruktur mit vielfach sensiblen Daten die **Verknüpfung der Datenbestände** elementar. Sie steigert das analytische Potential von Daten und ermöglicht innovative Auswertungen. So wurden in der Vergangenheit Daten aus dem Sozioökonomischen Panel mit Daten aus der Rentenversicherung und der Arbeitsagentur oder Daten aus Befragungen von Flüchtlingen mit Informationen des Bundesamtes für Migration und Flüchtlinge verknüpft.²² Auch kleinräumige Geodaten (z.B. aus der amtlichen Lärmkartierung) erlauben zusammen mit Unternehmens- und Befragungsdaten, die räumlichen Einflussfaktoren auf soziales Handeln besser zu verstehen.²³ Vor allem wenn Daten über Individuen aus verschiedenen Quellen verknüpft werden sollen, sind die rechtlichen Vorgaben derzeit sehr restriktiv. Technische und organisatorische Maßnahmen zum Schutz vor Deanonymisierung reichen nicht aus, sondern es wird die informierte Einwilligung der Befragten gefordert, was für die Forschung eine häufig unüberwindbare Hürde bedeutet. Eine Vernetzung der Datenbestände und die Schaffung von Möglichkeiten für die Verknüpfung mit Geo-, Medien-, und amtlichen Daten braucht einen Rechtsrahmen, der auf die besonderen Bedarfe der wissenschaftlichen Forschung zugeschnitten ist und zugleich den Schutzerfordernissen der unmittelbar und mittelbar an der Forschung Beteiligten Rechnung trägt. Nicht die Verknüpfung von Daten ist das Problem, sondern die Motive bei der Verwendung verknüpfter Daten.

Ähnliche Herausforderungen ergeben sich auch aus der Zusammenarbeit der NFDI-Konsortien. Die NFDI wird dabei helfen, Kompetenzen aufzubauen und gemeinsame Lösungen an den Stellen zu finden, an denen sich verschiedene Wissenschaftsbereiche mit ähnlichen Problemen beschäftigen. Die Herausforderung dabei wird sein, diese Gemeinsamkeiten zu entdecken und zu beginnen, in diesen wichtigen Fragen eine gemeinsame Sprache zu sprechen. Dazu wird sich KonsortSWD auch in die Diskussion zu den NFDI-weiten, sogenannten, Cross-Cutting Topics einbringen. In jedem Fall wird die Beantwortung vieler Fragestellungen aus den Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften von der Nachnutzung und Verknüpfung mit Daten aus anderen Disziplinen, wie beispielsweise Klimadaten oder Satellitendaten, profitieren.

Eines der aus unserer Sicht wichtigsten Cross-Cutting Topics ist die Forschungsethik bzw., damit zusammenhängend, die **Forschungsdatenethik**. Forschungsethik formu-

²²Manuel Siegert und Nina Rother, "Die IAB-BAMF-SOEP-Befragung von Geflüchteten" in *Forschungsdaten für die Kinder- und Jugendhilfe*, Maik-Carsten Begemann und Klaus Birkelbach, ed. (Wiesbaden: Springer VS, 2019), 187-209.

²³Stefan Schweers, Katharina Kinder-Kurlanda, Stefan Müller und Pascal Siegers, "Conceptualizing a spatial data infrastructure for the social sciences: An example from Germany," *Journal of Map & Geography Libraries* 12, no. 1 (2016): 100-126, doi: [10.1080/15420353.2015.1100152](https://doi.org/10.1080/15420353.2015.1100152).

liert Grundprinzipien eines für die empirische Forschung angemessenen Handelns im Hinblick auf die Beziehungen zwischen Forschenden und Teilnehmenden an Studien als auch hinsichtlich der Transparenz der Forschung und der Abschätzung von Folgewirkungen eines Forschungsvorhabens. Zentrales Ziel forschungsethischer Normen und Regelungen ist die Anleitung zu einem Handeln, das Schädigungen von Einzelnen, Gruppen, Organisationen oder ganzen Gesellschaften durch Forschung möglichst vermeidet und *gleichzeitig* Forschungsfreiheit ermöglicht. Dies betrifft insbesondere den Schutz von vulnerablen Gruppen wie beispielsweise von Opfern politischer Repression. Neue, weitreichende Herausforderungen für forschungs(daten)ethische Fragen ergeben sich zudem durch technologische Entwicklungen im Bereich der Datenerhebung und -auswertung, durch Anforderungen von Forschungsförderern und Verlagen an die Zugänglichkeit von Forschungsdaten sowie durch die zunehmende Komplexität von Forschungsdesigns aufgrund der angesprochenen Verknüpfung unterschiedlichster Datensätze.

Derzeit arbeiten viele lokale und dezentrale Ethikkommissionen weitgehend unvernetzt. Zugleich fehlen Forschenden vielfach - dies ergaben Workshops mit den Communities - adäquate Beratungsmöglichkeiten²⁴. Lösungen aus den medizinischen und psychologischen Communities sind hier nur bedingt übertragbar. Im Bereich der Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften setzt sich der RatSWD daher für ein gemeinsames Verständnis forschungsethischer Grundsätze und Prüfverfahren ein²⁵ und führt derzeit systematische Bedarfsermittlungen durch, auf deren Basis konkrete Empfehlungen und Handreichungen für den Aufbau und die Arbeit von lokalen Ethikkommissionen erarbeitet werden sollen. Die Herausforderung für die Zukunft ist eine forschungs(daten) ethische Infrastruktur, die den Austausch zwischen den lokalen Ethikkommissionen unterstützt, Forschende in ihrem Alltag beratend zur Seite steht und zugleich fachwissenschaftlich-disziplinären Anforderungen gerecht wird.

Auch und gerade angesichts dieser Herausforderungen freuen wir uns, zusammen mit den anderen Konsortien eine leistungsfähige Forschungsdateninfrastruktur in der NFDI aufzubauen, die qualitativ hochwertige Daten so bereitstellt, dass sie den größtmöglichen Mehrwert für unser Verständnis der Gesellschaft und ihrer Probleme ermöglichen.

²⁴Hollstein, Betina und Jörg Strübing, *Archivierung und Zugang zu Qualitativen Daten*. RatSWD Working Papers 267/2018. (Berlin: Rat für Sozial- und Wirtschaftsdaten, 2018), 1-14, doi: [10.17620/02671.35](https://doi.org/10.17620/02671.35).

²⁵Rat für Sozial- und Wirtschaftsdaten (RatSWD), *Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften* (Berlin: Rat für Sozial- und Wirtschaftsdaten, 2017), doi: [10.17620/02671.1](https://doi.org/10.17620/02671.1)